

# ビジネス文書を対象とした大規模言語モデルを用いた 読み手にストレスを与える文章の検出

松永剛之進<sup>1</sup> 野島瞳<sup>2</sup> 森下洋平<sup>2</sup> 青木輝勝<sup>1,3</sup>

<sup>1</sup>東京工科大学大学院 バイオ・情報メディア研究科

<sup>2</sup>一般社団法人 ユニバーサルコミュニケーションデザイン協会

<sup>3</sup>東京工科大学 コンピュータサイエンス学部

g212302822@edu.teu.ac.jp, {h-nojima, ymorishita}@ucda.jp

aokitms@stf.teu.ac.jp

## 概要

本研究は、大規模言語モデルを用いて、ビジネス文書中の読み手にストレスを与える文章を検出する手法を提案する。BERT(Bidirectional Encoder Representations from Transformers)[1]を用いた多クラス分類モデルにより、過剰な丁寧さ、命令的な表現、情報の過不足、責任回避的な表現を検出し、文章改善の指針を提供する。独自のデータセットを作成し、分類タスクに適した学習を行うことで、従来手法では検出が困難だった微妙な文体の違いを捉えることを目指す。実験により、定義したラベルは文章の内容に関わらず、ストレスの要因となる表現を検出可能であることを示す。

## 1 はじめに

情報過多の現代においては、文章を読む際に伝わりにくい表現や、責任の所在を不透明にする表現が使用されていることで、特にビジネス関連の文書で読み手がストレスを感じてしまうことが少なくない。このような書き手の意図していない伝わり方を防ぐために、文書の校正を行う専門団体も存在する。しかし、専門家が手動で校正を行う場合、需要に対して供給が追いついていない現状がある。また、近年では生成 AI の普及が進んでいるが、多タスク向け Chatbot は、ビジネス文章の校正を依頼しても明確な基準がなく、判断が感覚に依存してしまう傾向がある。本研究ではこの課題を解決するために、大規模言語モデルを用いて、ビジネス文書に対して読み手にストレスを与える文章を検出する手法を提案する。独自に作成したデータセットをモデルの学習に使用し、ストレスを与える文章の検出に特化したシステムの実現を目指す。本研究は、文章品質の向上

と深層学習を用いた文章校正を通じて、より良いビジネス文書作成を支援することを目標とする。

## 2 研究目的

本研究の目的は、大規模言語モデルを用いた読み手にストレスを与える文章を検出する手法を提案することである。「情報の質」に着目し、読み手が文章を読んだ際に感じるストレスの要因を定義する。本研究では、文章の内容そのものではなく、文体によって変化する伝わり方に注目して評価を行う。データセットは生成 AI を用いて、定義したラベルに基づいて作成し、従来のルールベースでは検出が困難であった文体の特徴を捉えることを目指す。

システムの主な利用者は、消費者向けに文章を作成する企業を想定し、検出対象とする文章は、保険の案内などの個人顧客向けビジネス文書とする。

## 3 文章 DC9 ヒューリスティック評価

表 1 評価項目

評価項目	指摘項目
表記	表記の統一性や親しみやすさ
語彙	専門用語の使用や不快語・否定語の有無
敬語	敬語の誤用や乱用の有無
文法	文法的な誤りの有無
文意	伝えたい内容や起こしてほしい行動の明確さ
文の構造	文の長さや構文上の問題
文章構造	情報の提示順序や構造化の適切さ
情報の質	不要な情報や曖昧な情報の有無 必要な情報の欠如
全体的な難易度	読みやすさや全体的な難易度の適切さ

文章 DC9 ヒューリスティック評価[2]は、文章の「わかりにくさ」を定量的に評価し、改善するために開発された評価手法である。この評価手法では、文章の特性を 9 つの評価項目に基づいて評価し、それぞれの項目ごとに具体的な指摘ポイントを設けている。表 1 は、文章 DC9 ヒューリスティック評価における評価項目と指摘項目である。この評価手法を用いて、専門家が文章の質を評価し、改善案を提案する。

下記の図 1 は「情報の質」の項目を詳細に示したものである。本研究では、この「情報の質」に着目し、読み手がストレスを感じる要因を分析する際の指標とする。

図 1 情報の質

情報の質：読み手に誤解とストレスを与えない
<ul style="list-style-type: none"> <li>曖昧な表現を避ける</li> <li>重複情報を避ける</li> <li>受け身の表現を避ける</li> <li>責任回避と誤解される文章を避ける</li> </ul>

## 4 提案手法

### 4.1 検出システム

本提案手法では、文章 DC9 ヒューリスティック評価における「情報の質」の項目を指標とし、ストレスを感じる要因となる表現を 5 つ定義した。この 5 つの表現に「適正な表現」を加えた 6 つの表現をラベルとして設定し、それぞれにクラスを割り当てる（表 2 参照）。また、表 3 にラベルの概要を記載する。

表 2 クラス

Class	Label
0	適正な表現
1	過剰に丁寧な表現
2	命令的な表現
3	情報が重複している表現
4	情報が不足している表現
5	責任回避的な表現

表 3 ラベル概要

クラス	概要
0	1~5 のクラス以外の適正な表現がなされている文章

1	必要以上の枕詞を使用し、丁寧すぎて冗長に感じられる表現
2	直接的すぎる表現や威圧的に感じられる文章
3	1 文内に重複や必要以上の情報が含まれている文章
4	指している内容が不明確で、重要な情報がわかりにくい表現
5	責任の所在が曖昧で、信頼性を損なう表現

ストレス要因を検出するモデル構築のために BERT を使い、多クラス分類タスクで学習を行う。提案モデルは、文章を一文ずつ解析し、分類結果に基づきストレスを感じる要因となる表現を検出する。このシステムは検出された文章とその理由（ストレス要因）をユーザに知らせ、文章改善を支援する。

### 4.2 データセット

ファインチューニング用データセットとして、金融関連の文章 100 文を収集した。本研究では句点 1 つのみ含む文章を 1 文と定義する。収集したデータを ChatGPT (Generative Pre-trained Transformer) [3] で各クラスの表現に言い換えた。1 データにつき 1 クラス 10 文を生成し、合計 6,000 文のデータセットを作成した（各クラス 1,000 文）。表 4 は作成したデータセットの一例であり、表内の原文は収集した文章を指す。原文を生成 AI によってクラス 0 から 5 のそれぞれの特徴を持たせ、データを生成した。データセットは各クラスの特徴をモデルが学習し、従来手法では難しかった微妙な文体差を検出できるように設計されている。検出時にはラベル情報を提示することで、ユーザが文章の問題点を特定し、改善に役立てられることが可能である。

図 2 に各クラスのトークン数(文字数)の分布を示す。データセットは、訓練用に 6 割、検証用とテスト用にそれぞれ 2 割ずつに分割する。

図 2 クラスごとのトークン数の分布

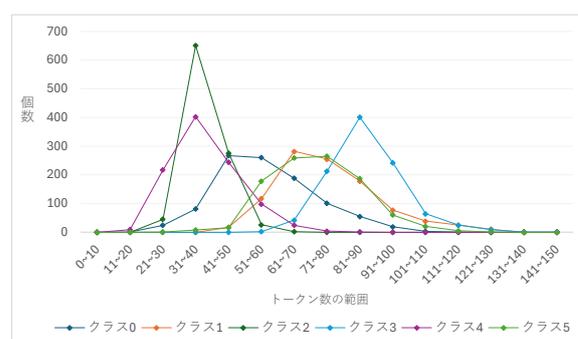


表 4 データセット例

Class	Data
原文	死亡・高度障害保険金に関しては,(災害)高度障害保険金をお支払いしたときは所定の高度障害状態になったときにさかのぼって契約は消滅し,その後に死亡された場合でも(災害)死亡保険金はお支払いいたしません.
0	死亡・高度障害保険金については,災害高度障害保険金をお支払いした際には,所定の高度障害状態にさかのぼって契約は消滅し,その後に死亡された場合でも災害死亡保険金はお支払いいたしかねます.
1	死亡・高度障害保険金に関して,誠に恐縮ですが,(災害)高度障害保険金をお支払いさせていただいた際には,所定の高度障害状態にさかのぼって契約が消滅いたします.その後,万が一死亡された場合でも,(災害)死亡保険金はお支払い致しかねます.
2	死亡・高度障害保険金は,(災害)高度障害保険金を支払った時点で契約を即時消滅とし,その後に死亡しても保険金は一切支払われない.
3	死亡保険金および高度障害保険金に関しては,災害により所定の高度障害状態に至った場合に高度障害保険金をお支払いすることが決定した際,その高度障害状態が発生した時点にさかのぼって保険契約は自動的に消滅いたしますので,仮にその後被保険者が死亡されたとしても,災害死亡保険金をお支払いすることはできず,契約がすでに終了しているため,保険金の支払い対象外となります.
4	死亡・高度障害保険金に関して,災害高度障害保険金をお支払いした際には,所定の高度障害状態になったとみなして消滅し,その後死亡された場合でも死亡保険金は控除されるとされています.
5	死亡・高度障害保険金に関しては,(災害)高度障害保険金をお支払いした時点で契約は消滅とさせていただきますので,その後の死亡に関しては(災害)死亡保険金のお支払いはご遠慮いただく形となります.

## 5 実験

### 5.1 実験概要

本実験では,大規模言語モデル BERT を用いた多クラス分類を実施する.使用するモデルは,東北大学 乾研究室が公開している日本語モデル「bert-base-japanese-v3」[4]である.訓練データと検証データは本研究において作成したデータセットを用い,モデルの追加学習を行った.モデルの評価には,テストデータを用いて正解率を測定し,分類性能を検証する.

実験では,データの偏りを防ぐため,学習およびテストにおいてデータをランダムに分割し,10回の試行を行った.表 5 にはモデルの学習に設定した主なハイパーパラメータや関数を示す.

表 5 モデルに用いたハイパーパラメータ

パラメータ	値
学習率	0.00001
バッチサイズ	64
最大エポック数	30
最大入力長	256

ドロップアウト率	0.5
損失関数	クロスエントロピー損失
重み減衰率	0.01
精度計算方法	重み付き平均 (weighted)
最適化手法	AdamW

### 5.2 実験結果

表 6 は各クラスにおけるモデルの試行回数 10 回の平均精度 (Avg Accuracy), 最高精度 (Max), 最低精度 (Min) を示している.全体の平均精度は 0.880 であり,各クラス間で一定のばらつきが見られた.

表 6 モデルの精度評価

Class	Avg Accuracy	Max	Min
0	0.754	0.865	0.675
1	0.979	1.000	0.935
2	0.996	1.000	0.985
3	0.879	0.960	0.800
4	0.785	0.940	0.605
5	0.891	0.925	0.845
Overall	0.880	0.899	0.863

表7は各クラスの正解データ(Ground Truth)と予測クラス(Predicted Class)の対応を示す混同行列である。クラス1とクラス2は高い精度を示し、クラス1は2000件中1957件、クラス2は1992件が正しく分類された。一方、クラス0は230件、クラス4は399件が他クラスに誤分類された。特にクラス0とクラス4の間での混同が多く見られた。クラス3は2000件中1757件が正しく分類され、クラス5は1781件が正解であったが、それぞれ一部が他クラスに誤分類される結果となった。全体として、クラスごとに分類精度に差があることが確認された。

表 7 各クラスの分類結果

		Predicted Class					Total	
		0	1	2	3	4		5
Ground Truth	0	1508	26	0	159	230	77	2000
	1	9	1957	0	33	0	1	2000
	2	2	0	1992	1	5	0	2000
	3	49	10	0	1757	0	184	2000
	4	399	0	11	1	1569	20	2000
	5	54	0	0	155	10	1781	2000
Total		2021	1993	2003	2106	1814	2063	12000

## 5.3 考察

実験では、BERT モデルによる多クラス分類の性能が示された。以下に各クラスの特徴と課題について考察する。

### 5.3.1 クラス1・クラス2の精度

クラス1とクラス2は非常に高い分類精度を達成しており、モデルがこれらのクラスに対する特徴を十分に学習できていることを示している。これにより、情報量が豊富かつ特徴が明確な表現では高い分類性能が期待できる。

### 5.3.2 クラス0・クラス4の課題

クラス0は230件、クラス4は399件が誤分類されていることから、これらのクラス間に類似点があり、区別が難しいと推測される。特にクラス0はクラス4に誤分類される傾向が強く、表現の曖昧さや情報不足などが影響している可能性がある。

### 5.3.3 クラス3の誤分類

クラス3は184件がクラス5に誤分類されており、情報過多な文章がより簡略化された表現に近いと誤認される場合がある。

### 5.3.4 改善案

誤分類を減らすため、データセットの増強やデータ前処理方法の改善、ラベル間の類似性を反映した損失関数の適用を検討することで精度向上が見込まれる。また、クラスごとに異なる特徴量を追加することにより、表現の違いをより正確に捉えられるモデル構築も求められる。

## 6 おわりに

本研究では、ビジネス文章を対象とした大規模言語モデルを用いた読み手にストレスを与える文章の検出システムを構築し、その有用性を実証した。提案手法により、多様な文体を分類し、文章品質を向上させるための具体的な改善案を提示できることを確認した。しかし、クラス間の類似性による誤分類が発生する課題も明らかになった。特に適正な表現と情報が不足している表現の区別は困難であり、さらなる精度向上のためには、データの多様性の確保が求められる。

今後は、異なる領域の文章データセットを活用した汎用性の評価や、ビジネスシーンにおける文章校正支援の実用化を目指す。これにより、生成AI時代におけるコミュニケーション品質の向上に寄与することができる。

## 参考文献

- [1] Devlin, J., Chang, M.-W., Lee, K., & Toutanova, K. (2019). BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. *NAACL-HLT*, 4171-4186.
- [2] UCDA. 文章 DC9 ヒューリスティック評価, [https://ucda.jp/solutions/text\\_dc9.html](https://ucda.jp/solutions/text_dc9.html)
- [3] OpenAI. (2024). ChatGPT[Large language model]. OpenAI. Available at: <https://www.openai.com/>
- [4] 東北大学乾研究室. bert-base-japanese, <https://onl.bz/VsMxmgV>